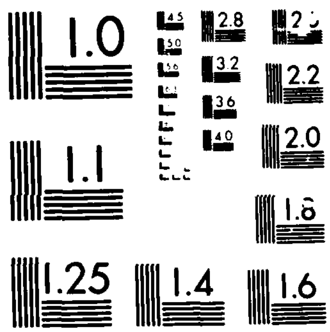
AD-A169 691 TWO-SAMPLE INFERENCE BASED ON ONE-SAMPLE WILCOXON SIGNED RANK STATISTICS..(U) PENNSYLVANIA STATE UNIV UNIVERSITY PARK DEPT OF STATISTICS M TABLEMAN ET AL. 1/1

UNCLASSIFIED JUN 86 TR-61 N00014-80-C-0741 F/G 12/1 NL


1.0
2.8
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

The Pennsylvania State University
# Department of Statistics
University Park, Pennsylvania

TECHNICAL REPORTS AND PREPRINTS

Number 61: June 1986

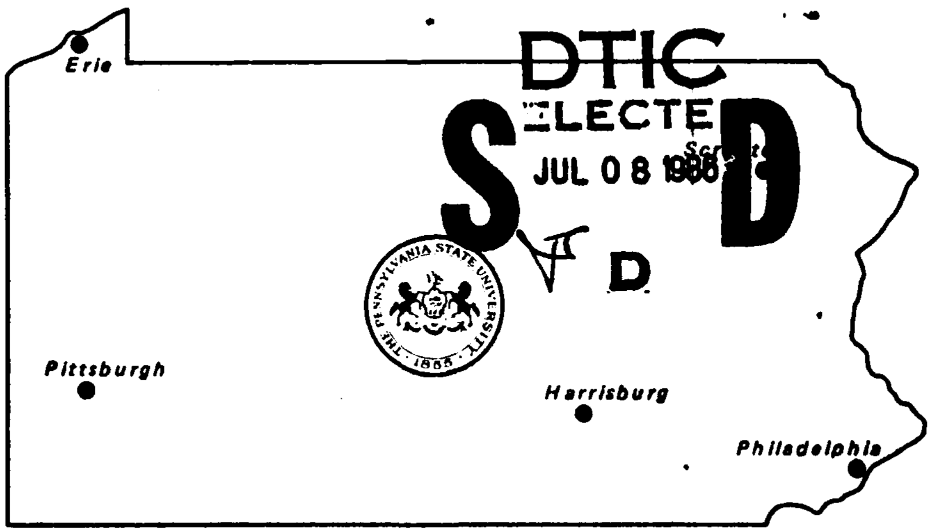
# TWO-SAMPLE INFERENCE BASED ON ONE-SAMPLE WILCOXON SIGNED RANK STATISTICS

Mara Tableman
Eidgenoessische Technische Hochschule

and

Thomas P. Hettmansperger*
The Pennsylvania State University





(12)

DEPARTMENT OF STATISTICS

The Pennsylvania State University
University Park, PA 16802 U.S.A.

TECHNICAL REPORTS AND PREPRINTS

Number 61: June 1986

# TWO-SAMPLE INFERENCE BASED ON ONE-SAMPLE WILCOXON SIGNED RANK STATISTICS

Mara Tableman
Eidgenoessische Technische Hochschule

and

Thomas P. Hettmansperger*
The Pennsylvania State University



*Research partially supported by ONR Contract N00014-80-C0741.

# Two-Sample Inference Based on One-Sample Wilcoxon Signed Rank Statistics

Mara Tableman
Seminar fur Statistik
Eidgenoessische Technische Hochschule
CH-8092 Zurich
Switzerland

Thomas P. Hettmansperger*
Department of Statistics
Pennsylvania State University
University Park, PA 16802

## Abstract

A two-sample test is studied which rejects the null hypothesis of equal population medians when two Wilcoxon distribution free confidence intervals are disjoint. A confidence interval for the difference in population medians is constructed by subtracting the endpoints of two one-sample confidence intervals. Two different ways to select the one-sample intervals are presented. A solution that specifies equal confidence coefficients for the one-sample intervals is recommended. All solutions are shown to have the same asymptotic (Pitman) efficiency as the Mann-Whitney two-sample test.

Key Words and Phrases: Mann-Whitney-Wilcoxon test, nonparametric test, sign test, nonparametric confidence intervals, notched box plots.

*Research partially supported by ONR Contract N00014-80-C0741.

## 1. Introduction

The purpose of this paper is to discuss a two-sample test and confidence interval based on the comparison of one-sample confidence intervals. This approach is appealing because, along with the two-sample inference, a statistical description of the individual samples is provided and a graphical representation of the two intervals provides an effective presentation of the inference. (See Figure 1). The test rejects the null hypothesis of equal population medians if the one-sample intervals are disjoint. The confidence interval for the difference in population medians is constructed by subtracting the endpoints of one interval from the opposite endpoints of the other.

McGill, Tukey and Larsen (1978) discuss a similar idea when presenting notched box plots as a way of displaying relevant information about a population. The notch in their box is a confidence interval of the form $M \pm 1.7SE$ for the population median, where M is the sample median and SE is a sample estimate of the asymptotic standard error of the sample median when sampling from a normal population. The factor 1.7 was empirically chosen to produce a two-sample test with a two-sided level of roughly 5% for several distributions. It is difficult to interpret the individual intervals (notches) and "level" of the test since they are based on such rough approximations. McKean and Schrader (1983), in a simulation study, found the confidence coefficient for this procedure to be highly unstable for small to moderate sample sizes, even at the normal distribution. Hettmansperger (1984) suggested notching the box plots with sign-intervals. These intervals have order statistics from the sample as endpoints and are obtained by inverting the acceptance region of a two-sided sign test. These intervals are distribution free, and have a well-defined asymptotic distribution theory that does not require normality or symmetry assumptions. He also showed that this procedure has the same asymptotic efficiency as Mood's (1950) median test which corresponds to a particular choice of sign-intervals. Tableman (1984) developed the small sample distribution theory for the comparison of sign-intervals so that one can easily obtain the exact level and confidence coefficient of the inference.

For the case of symmetric populations, we propose notching the box plots with one-sample Wilcoxon-intervals which are defined in Section 2. It is shown that these intervals are distribution free, and exact confidence coefficients can be obtained from a table of the Wilcoxon (1945) signed-rank distribution. Several large sample results concerning its endpoints are presented. In Section 3, the two-sample inference is developed and discussed. It is shown that the suggested procedures produce two-sample tests that have the same Pitman efficiency as the Mann-Whitney-Wilcoxon (1945, 1947) two-sample test, when sampling from symmetric distributions. Hence, the proposed procedure can be much more efficient than the procedure based on sign-intervals. Two choices of confidence coefficients for the two intervals are discussed and an example is presented. In the final section, we discuss the effects of asymmetry on the two-sample inference. Although in simulation studies we have not found the size of the test to be effected, a reduction in power (Pitman efficiency) of the proposed test can result. Hence, with strong asymmetry, it may no longer be a competitor to the Mann-Whitney-Wilcoxon test. However, in many models it may not be unreasonable to assume symmetry or approximate symmetry, especially after a transformation has been applied.

## 2. The Wilcoxon-Interval

Assumptions 2.1:

(i) Suppose $X_1,\ldots,X_m$ is a random sample of size m with cumulative distribution function (cdf) $F(x-\theta)$, where $\theta$ is the unique median, F is continuous, and has density f, symmetric about 0, and which satisfies $\int f^2(x)dx<\infty$.

(ii) Let $G_\theta(t)$ denote the cdf of $(X_1+X_2)/2$. Let $\xi_p$ denote the pth quantile of $F(x-\theta)$, $0<p<1$. Suppose $G_\theta$ is twice differentiable in a neighborhood of $\xi_p$, with $G_\theta'=g_\theta$ positive and $G_\theta''$ bounded in the neighborhood.

Remark 2.1. If f is symmetric, absolutely continuous, and has finite Fisher's information $\int_{-\infty}^{\infty}(f'(x)/f(x))^2f(x)dx < \infty$, then the Assumptions 2.1 concerning F and $G_\theta$ are satisfied. For a proof, see Aubuchon (1982, p.19).

The Wilcoxon confidence interval for $\theta$, the true population median, is derived by inverting the acceptance region of a size $\alpha=2P_\theta(T_m(\theta)<d(m)) = 2P_0(T_m(0)<d(m))$ Wilcoxon signed rank test where

$$T_m(\theta) = \sum_{i=1}^{m} R(|X_i-\theta|)\cdot I\{X_i>\theta\} = \sum_{1\le i\le j\le m}\sum I\,\{(X_i+X_j)/2 > \theta\} \tag{2.1}$$

with $I\{A\}$ denoting the indicator function of the event A and $R(|X_i-\theta|)$ representing the rank of $|X_i-\theta|$ among the ordered $|X_j-\theta|$, $j=1,\ldots,m$. The $m(m+1)/2$ averages $(X_i+X_j)/2$, $1\le i\le j\le m$, are referred to as the Walsh averages.

Let $M = m(m+1)/2$ and let

$$Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(M)} \tag{2.2}$$

represent the ordered Walsh averages. In view of the second equality in (2.1),

$$\begin{aligned} &d(m) \leq T_m(0) \leq m(m+1)/2 - d(m) \quad \text{if and only if} \\ &[Z_{(d(m))}, Z_{(M-d(m)+1)}) \text{ contains } \theta. \end{aligned} \tag{2.3}$$

Since we assume F is continuous, we take the Wilcoxon-interval for $\theta$ to be the closed interval

$$[Z_{(d(m))}, Z_{(M-d(m)+1)}]. \tag{2.4}$$

The d(m)-value will be called a depth since it specifies how deep into the ordered Walsh averages the endpoints lie. It is clear that this interval has exact confidence coefficient $\gamma=1-\alpha$ and is distribution free (since the Wilcoxon signed rank test statistic is).

Define,

$$G_m(x) = \sum_{1\leq i\leq j\leq m}\sum I\,\{(X_i+X_j)/2\leq x\}/\binom{m}{2} \tag{2.5}$$

and note that

$$\begin{aligned} \tilde{G}_m(x) &= 2T_m(\theta)/m(m+1) \\ &= G_m(x) + 0(m^{-1}). \end{aligned} \tag{2.6}$$

Now a slight adaptation of Lemma 4.2 of Geertsema (1970) yields the following lemma:

<u>Lemma 2.1.</u> Under Assumptions 2.1 (with $\xi_p=\theta$), if $\{k_m\}$ is any sequence of integers satisfying $2k_m/m(m+1) = 1/2 + O(m^{-1/2}\log m)$ as $m \to \infty$, then wp1

$$Z_{(k_m)} = \theta + [2k_m/m(m+1) - \tilde{G}_m(\theta)]/G_\theta'(\theta) + O(m^{-3/4}\log m) \qquad (2.7)$$

where $\tilde{G}_m$ is defined in (2.6) and $Z_{(k_m)}$ is defined in (2.2).

We have at once the following almost sure representation of the endpoints of the Wilcoxon-interval. The constant $k_m$ is suggested by the normal approximation used with the Wilcoxon statistic.

<u>Theorem 2.1.</u> Under Assumptions 2.1, wp1

$$Z_{(d(m))} = \theta - k_\alpha/((12m)^{1/2}\int f^2(x)dx)-[\tilde{G}_m(\theta)-1/2]/(2\int f^2(x)dx)+o(m^{-1/2})$$

$$(2.8)$$

$$Z_{(M_m-d(m)+1)} = \theta + k_\alpha/((12m)^{1/2}\int f^2(x)dx)-[\tilde{G}_m(\theta)-1/2]/(2\int f^2(x)dx)+o(m^{-1/2})$$

where $d(m) = m(m+1)/4 + .5 - k_\alpha(m(m+1)(2m+1)/24)^{1/2}$.

<u>Proof</u>: We prove the theorem for the lower endpoint $Z_{(d(m))}$. A similar argument holds for the upper endpoint.

It is easily shown that $d(m)/(m(m+1)/2) = 1/2-k_\alpha/(3m)^{1/2}+o(m^{-1/2}) = 1/2 + o(m^{-1/2}\log m)$. Now, $G_\theta'(\theta) = 2\int f^2(x)dx$, and observe that $O(m^{-3/4}\log m) = o(m^{-1/2})$. The theorem follows immediately from the preceding lemma.

<u>Remark 2.2.</u> Under Assumption 2.1(ii), Geertsema's result and hence Theorem 2.1 can be easily extended to include any quantile $\xi_p$, $0<p<1$.

Let $L = Z_{(d(m))}$ and $U = Z_{(M-d(m)+1)}$, the lower and upper endpoints of the Wilcoxon-interval (2.4). As a corollary to Theorem 2.1, we have the following:

<u>Corollary.</u>

As $m\to\infty$,

(i) $m^{1/2}(L-\theta) \xrightarrow{D} Z \sim \eta(-k_\alpha/12^{1/2}\int f^2(x)dx,\ 1/12(\int f^2(x)dx)^2)$

where $\eta(\mu,\sigma^2)$ denotes the normal cdf with parameters $\mu$ and $\sigma^2$.

(ii) $m^{1/2}(U-L) \to k_\alpha/(3^{1/2}\int f^2(x)dx)$ wp1. (2.9)

(iii) $P(\theta<L) \to \Phi(-k_\alpha) = \alpha$, where $\Phi(\cdot)$ is the standard normal cdf.

<u>Proof:</u> To prove part (i), we use the fact that $G_m(\cdot)$ (2.5) is a U-statistic. By a theorem due to Hoeffding (Serfling; 1980, p.192), $m^{1/2}(G_m(\theta)-1/2) \xrightarrow{D} \eta(0,1/3)$. Slutsky's Theorem along with (2.6) implies $m^{1/2}(\tilde{G}_m(\theta)-1/2) \xrightarrow{D}$

$\eta(0,1/3)$. Now, (i) follows at once from (2.8) and Slutsky's theorem. Part (ii) is immediate from (2.8). In part (iii), write $P(\theta<L)$ as $P(0<m^{1/2}(L-\theta))$ and apply part (i).

Hence, if the Wilcoxon-interval (2.4) is defined by depths

$$d(m) = m(m+1)/4+.5-k_\alpha(m(m+1)(2m+1)/24)^{1/2}, \tag{2.10}$$

the approximate confidence coefficient is $\gamma = 1-2\Phi(-k_\alpha) = 1-2\alpha$.

## 3. The Two-Sample Inference

Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ represent independent random samples from $F(x-\theta_x)$ and $F(y-\theta_y)$ respectively where

(i) $\theta_x$ and $\theta_y$ are the unique medians of the X and Y populations

(ii) F satisfies the Assumption 2.1

(iii) $G_{\theta_x}(t_x) = P((X_1+X_2)/2\le t_x)$ and $G_{\theta_y}(t_y) = P((Y_1+Y_2)/2\le t_y)$ satisfy the Assumption 2.1.

Further, with $M = m(m+1)/2$ and $N = n(n+1)/2$, let

$$[L_x,U_x] = [Z_{(d(m))},Z_{(M-d(m)+1)}],$$

and (3.1)

$$[L_y,U_y] = [Z_{(d(n))},Z_{(N-d(n)+1)}]$$

represent Wilcoxon-intervals (2.4) constructed on the X and Y samples, respectively, where

(i) the depths d(m) and d(n) are defined as in Theorem 2.1, however, with $k_x$ and $k_y$ replacing the $k_\alpha$, and

(ii) $\gamma_x = 1-2P(T_m(0)<d(m))$ and $\gamma_y = 1-2P(T_n(0)<d(n))$ denote the respective confidence coefficients.

To test

$$H_0: \Delta = \theta_y - \theta_x = 0 \text{ vs. } H_A: \Delta \neq 0$$

we reject $H_0$ if the two Wilcoxon-intervals are disjoint. That is,

$$\text{if } U_x < L_y \text{ or } U_y < L_x. \tag{3.2}$$

We wish to pick the two Wilcoxon-intervals so that if the intervals are disjoint we reject $H_0$: $\Delta=0$ in favor of $H_A$: $\Delta\neq0$ with a specified significance level $\alpha_c$.

Recall from the Corollary to Theorem 2.1, part (iii), that the approximate confidence coefficients are $\gamma_x \doteq 1-2\Phi(-k_x)$ and $\gamma_y \doteq 1-2\Phi(-k_y)$, respectively. The following theorem relates $\alpha_c$, the comparison rate, to $k_x$ and $k_y$ which determine $\gamma_x$ and $\gamma_y$. It therefore indicates how to pick $\gamma_x$ and $\gamma_y$ so that we achieve $\alpha_c$ (or $\gamma_c=1-\alpha_c$) at least in an approximate sense.

<u>Theorem 3.1.</u> Suppose that $m,n\to\infty$ so that $m/(m+n)\to\lambda$, $0<\lambda<1$. Then under $H_0$: $\Delta=0$, with the decision to reject given by (3.2),

$$\alpha_c = P(U_x<L_y) + P(U_y<L_x) \to 2\Phi(-(1-\lambda)^{1/2}k_x-\lambda^{1/2}k_y), \tag{3.3}$$

where $\Phi(\cdot)$ is the standard normal cdf.

The proof is the same as that of corollary to Theorem 2.1 of Hettmansperger (1984).

Define $k_c$ by $\alpha_c=2\Phi(-k_c)$. Then from (3.3) $k_x$ and $k_y$ must satisfy the condition

$$k_c = (1-\lambda)^{1/2}k_x+\lambda^{1/2}k_y. \tag{3.4}$$

Clearly, there are infinite number of choices for $k_x$ and $k_y$, hence for $\gamma_x$ and $\gamma_y$. We will discuss two of these choices later.

The test can be based on a confidence interval for $\Delta=\theta_y-\theta_x$. An equivalent decision is: Reject $H_o$: $\Delta=0$ in favor of $H_A$: $\Delta\neq0$ if

$$0 \text{ is not contained in } [L_y-U_x,\ U_y-L_x]. \tag{3.5}$$

Hence, if $\alpha_c$ is the significance level of the test, then the interval in (3.5) has confidence coefficient $\gamma_c=1-\alpha_c$. We therefore take the two-sample confidence interval for $\Delta=\theta_y-\theta_x$ to be

$$[L_y-U_x,U_y-L_x] = [Z_{(d(n))}-Z_{(M-d(m)+1)},Z_{(N-d(n)+1)}-Z_{(d(m))}]. \tag{3.6}$$

This interval has approximate confidence coefficient $\gamma_c=1-2\Phi(-k_c)$ where $k_c$ is defined in (3.4).

<u>Remark 3.1.</u> The present procedure is based on the ordered dependent Walsh averages $\{(X_i+X_j)/2;\ 1\le i\le j\le m\}$ and $\{(Y_i+Y_j)/2;\ 1\le i\le j\le n\}$. Hence, we are unable to formulate the procedure in terms of statistics whose distributions are tractable. We must rely on the normal approximation to the size $\alpha_c$. For a second-order approximation based on Edgeworth expansions see Tableman (1984).

The following theorem which is similar to Theorem 2.3 of Hettmansperger (1984), gives the asymptotic length of the two-sample confidence interval

(3.6) which is a measure of its efficiency and hence of the test from which it is derived. The asymptotic length does not depend on $k_x$ or $k_y$ but rather on $k_c$ as defined in (3.4).

Theorem 3.2. Suppose $m,n \to \infty$ so that $m/(m+n) \to \lambda$, $0<\lambda<1$. Let $\Lambda(TW)$ denote the length of the two-sample confidence interval given in (3.6). Then, wp1

$$(m+n)^{1/2}\Lambda(TW) \to k_c/(3\lambda(1-\lambda))^{1/2}\int f^2(x)dx. \tag{3.7}$$

We note that not only do all the intervals (obtained through the defining relationship (3.4)) have the same asymptotic length, but $N^{1/2}\Lambda(TW)/(2k_c)$ converges to $1/((12\lambda(1-\lambda))^{1/2}\int f^2(x)dx$ wp1. This limit is the reciprocal of the Pitman efficacy of the Mann-Whitney-Wilcoxon two-sample test; see Hettmansperger (1984, p.163). Hence, the tests based on the Wilcoxon-intervals have the same efficiency properties as the Mann-Whitney-Wilcoxon test.

As mentioned previously, there are an infinite number of ways to choose the intervals so that the two sample test has asymptotic size $\alpha$. For the case of sign-intervals Hettmansperger (1984) considered restricting the two intervals to having the same confidence coefficient or having the same asymptotic length. The same solution applies to the Wilcoxon-intervals. In particular, we recommend the equal confidence coefficients solution because, as indicated in Table 1 of Hettmansperger (1984), the common value of the confidence coefficients does not change much as the ratio of sample sizes varies between 1 and 3. Other solutions are not as stable.

Hence, we recommend using two 84% Wilcoxon-intervals for a 5% two-sided, two-sample test or two 75% Wilcoxon-intervals for a 10% two sided, two-sample test. The later case provides for a 5% one-sided, two-sample test.

Example. We illustrate the ideas on a small data set. Suppose we wish to compare the growth of mushrooms with and without vitamin B1 in their diets. The data given in Table 1 has been altered for illustration purposes from that given by Lehmann (1975, p.47). The measurement is weight gain in milligrams at the end of an observation period.

- Table 1 about here -

Let $\Delta=\theta_y-\theta_x$, the difference between the treatment (B1) and control (NO B1) population medians. We assume the weight gain populations are symmetric. To test $H_0$: $\Delta=0$ vs $H_A$: $\Delta\neq0$, we could use the Mann-Whitney-Wilcoxon rank sum test. Using the MANN command in Minitab (see Ryan, Joiner and Ryan (1985)), we find the sum of the ranks (breaking ties by averaging) of the treatment group is 141.5 with a two sided p-value of .0065. Hence we reject $H_0$: $\Delta=0$ at $\alpha<.01$.

A more informative analysis is based on the comparison of two 84% Wilcoxon confidence intervals for $\theta_x$ and $\theta_y$. Figure 1 provides the notched box plots for the two groups. Comparing the 85%-Wilcoxon notches leads to rejection of $H_0$: $\Delta=0$ at the 5% level since they are disjoint.

- Figure 1 about here -

Further, we have the intervals and estimates for $\theta_y$ and $\theta_x$ as $[L_y,U_y]$ = [23.0,28.5], $\hat{\theta}_y$=25.5 and $[L_x,U_x]$ = [15.0,20.5], $\hat{\theta}_x$=18.0, respectively. The estimates $\hat{\theta}_x$ and $\hat{\theta}_y$ are the Hodges-Lehmann (1963) estimates computed as the median of the Walsh averages. In addition, a 95% confidence interval for $\Delta$ is given by $[L_y-U_x,U_y-L_x]$ = [2.5,13.5] with $\hat{\Delta} = \hat{\theta}_y-\hat{\theta}_x = 7.5$. The figure along with the intervals and estimates are available using the BOXPLOT and WINT commands in Minitab.

Note that the Hodges-Lehmann estimate and 95% confidence interval for $\Delta$ derived from the Mann-Whitney statistic are $\hat{\Delta} = 8.0$ and [3.0,13.0], respectively. These differ slightly from the Wilcoxon-interval comparison.

Eighty five percent sign-notches are not available for the given sample size. The closest confidence coefficient to 85 percent is 89 percent and yields [22.0,29.0] and [14.0,23.0] for B1 and NO B1 groups, respectively. These intervals overlap and would fail to reject $H_o$: $\Delta$=0 at some level less than .05. If the distributions are not too heavy tailed then the Wilcoxon notches and the Mann-Whitney-Wilcoxon test are more efficient than the sign-notches and this may result in a failure of the sign-notches to detect a significant difference when the other methods will detect it.

## 4. The Effects of Asymmetry on the Two-Sample Inference

As remarked in Section 2, Geertsema's result can be extended to include any quantile $\xi_p$, $0<p<1$. We state here, without proof, the form most useful for our discussion. We note that the result does not rely on the symmetry of F.

Theorem 4.1. Under Assumptions 2.1 (i), (ii) (but with symmetry on F dropped), if $\{k_m\}$ is any sequence of integers satisfying $2k_m/m(m+1) = G_\theta(\xi_p)+k/m^{1/2}+o(m^{-1/2})$ as $m\to\infty$, then wp1

$$Z_{(km)} = \xi_p+k/m^{1/2}G_\theta'(\xi_p) - [\tilde{G}_m(\xi_p)-G_\theta(\xi_p)]/G_\theta'(\xi_p) + o(m^{-1/2}).$$

Let $\theta^*$ denote the median of $G_\theta(\cdot)$. Note that under asymmetry, $\theta^*\neq\theta$. With the depth d(m) defined as in Theorem 2.1, we have $2d(m)/m(m+1) = 1/2-k/(3m)^{1/2}+o(m^{-1/2})$ as $m\to\infty$. But now, $1/2=G_\theta(\theta^*)$. Hence, Theorem 4.1 gives us the following almost sure representation of the endpoints of the one-sample Wilcoxon-interval. They differ from those given in Theorem 2.1.

$$L = \theta^*-k/(3m)^{1/2}G_\theta'(\theta^*)-[\tilde{G}_m(\theta^*)-1/2]/G_\theta'(\theta^*)+o(m^{-1/2})$$

and (4.1)

$$U = \theta^*+k/(3m)^{1/2}G_\theta'(\theta^*) -[\tilde{G}_m(\theta^*)-1/2]/G_\theta'(\theta^*)+o(m^{-1/2}).$$

Before proceeding to the two sample inference, we make the following remarks:

Remarks 4.1.

(i) When F is not symmetric, $\theta^*$ differs from $\theta$ by a constant a; that is, $\theta^*(\theta)=\theta+a$. This follows from the fact that $\theta^*$ is such that

$$G_\theta(\theta^*)=\int_{-\infty}^{\infty} F(2\theta^*-2\theta-x)f(x)=1/2.$$

Now, differentiate with respect to $\theta$ to obtain $d(\theta^*(\theta)/d\theta=1$. Hence, $G_\theta'(\theta^*) = 2\int_{-\infty}^{\infty} f(2a-x)f(x)dx$.

(ii) It follows from Theorem A in Serfling (1980, p.192) and from (2.6) that $m^{1/2}[\tilde{G}_m(\theta*)-1/2]/G_\theta'(\theta^*) \overset{D}{\to} Z\sim\eta[0;(\nu-1/4)/(\int_{-\infty}^{\infty} f(2a-x)f(x)dx)^2]$ as $m\to\infty$, where $\nu = \int_{-\infty}^{\infty} [F(2a-x)]^2 f(x)dx$.

We now return to the two-sample inference. Let $\theta_x^*$ and $\theta_y^*$ denote the medians of $G_{\theta x}(\cdot)$ and $G_{\theta y}(\cdot)$, respectively. Under the same shape assumption on the underlying distributions, $\theta_x^*=\theta_x+a$ and $\theta_y^*=\theta_y+a$. It is clear that testing $H_o$: $\theta_y=\theta_x$ is equivalent to testing $H_o$: $\theta_y^* =\theta_x^*$, and that $\Delta=\theta_y^*-\theta_x^*$. We are now in a position to approximate $\alpha_c$, the comparison rate of the test (3.2), and to determine the asymptotic length of the two-sample interval (3.6) for $\Delta$ when F is asymmetric. These results are stated in the next theorem and they follow immediately from (4.1) and Remarks 4.1.

Theorem 4.2. Suppose that $m,n\to\infty$ so that $m/(m+n)\to\lambda$, $0<\lambda<1$.

(i) Then under $H_o$: $\Delta=0$, with the decision to reject given by (3.2)

$$\alpha_c \to 2\Phi(-k_c/(12(\nu-1/4))^{1/2}).$$

(ii) Let $\Lambda(TW)$ denote the length of the two-sample confidence interval given in (3.6). Then, wp1,

$$(m+n)^{1/2}\Lambda(TW) \to k_c/[(3\lambda(1-\lambda))^{1/2}\int_{-\infty}^{\infty} f(2a-x)f(x)dx].$$

Remarks 4.2.

(i) Suppose $k_c$ is the critical value that provides a size-$\alpha_c$ two-sided test when $F(\cdot)$ is symmetric. Although $\nu=1/3$ when $F(\cdot)$ is symmetric, for asymmetric distributions we have $1/4 \le \nu \le 1/2$. We see from part (i) of Theorem 4.2 this can result, theoretically, in increasing the size. However, Tableman (1984) simulated the two-sample test (3.2) for samples of small to moderate sizes from chi-square distributions with 4, 18, 30 degrees of freedom and from the standard normal distribution. No evidence was found that asymmetry (in this case, right skewness) effects the size.

(ii) Part (ii) of Theorem 4.2 reveals that the main effect of asymmetry is the possibly significant reduction in power. This follows from the fact that for certain asymmetric densities $\int_{-\infty}^{\infty} f(2a-x)f(x)dx$ can be quite small. We also note that the Pitman efficiency is no longer that of the Mann-Whitney two-sample test. In this setting, the efficacy of that test remains $(12(1-\lambda)\lambda)^{1/2}\int_{-\infty}^{\infty} f^2(x)dx$, and hence may be much more efficient since $\int_{-\infty}^{\infty} f(2a-x)f(x)dx \le \int_{-\infty}^{\infty} f^2(x)dx$.

## REFERENCES

Aubuchon, J. C. (1982). Rank Tests in the Linear Model: Asymmetric Errors. Unpublished Ph.D. dissertation, The Pennsylvania State University, University Park.

Geertsema, J. C. (1970). Sequential confidence intervals based on rank tests. Ann. Math. Statist. 41, 1016-1026.

Hettmansperger, T. P. (1984). Two-sample inference based on one-sample sign statistics. J. R. Statist. Soc. Ser. C. 33, 45-51.

Hettmansperger, T. P. (1984). Statistical Inference Based On Ranks. John Wiley, New York.

Hodges, J. L., Jr. and Lehmann, F. L. (1963). "Estimates of Location Based on Ranks," Ann. Math. Statist. 34, 598-611.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, San Francisco.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50-60.

McGill, R., Tukey, J. W. and Larsen, W. A. (1978). Variations of box plots. The American Statistician 32, 12-16.

McKean, J. W. and Schrader, R. M. (1983). A comparison of methods for studentizing the sample median. Tech. Rpt. No. 68, Dept. of Math., Western Michigan University, Kalamazoo, Michigan.

Mood, A. M. (1950). Introduction to the Theory of Statistics. McGraw-Hill, New York.

Ryan, B. F., Joiner, B. L., Ryan, T. A., Jr. (1985). Minitab Handbook, 2nd Ed., Duxbury, Boston.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.

Tableman, M. (1984). Two-Sample Procedures Based on One-Sample Linear Signed Rank Statistics. Unpublished Ph.D. dissertation, The Pennsylvania State University, University Park.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics 1, 80-83.

Table 1

Weight Gain

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| X: | NO B1 | 12 | 13 | 14 | 14 | 17 | 18 | 18 | 23 | 24 | 25 |
| Y: | B1 | 19 | 20 | 22 | 22 | 24 | 27 | 27 | 29 | 34 | 35 |

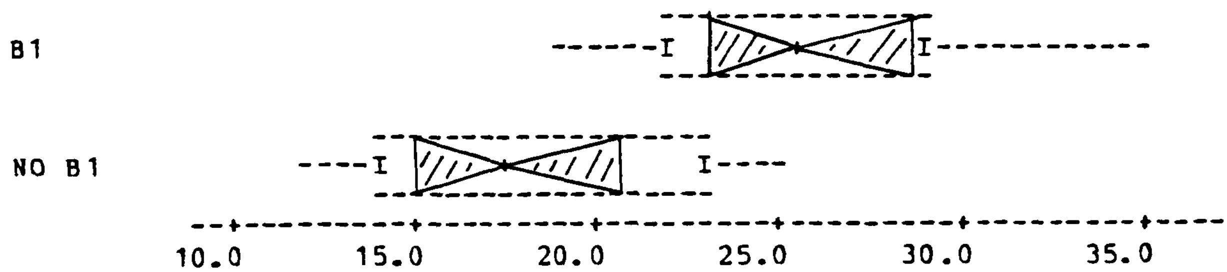



Figure 1. Wilcoxon Notched Boxplots for Mushroom Data.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
61

2. GOVT ACCESSION NO.
ADA 169691

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Two-Sample Inference Based on One-Sample Wilcoxon Signed Rank Statistics

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER


7. AUTHOR(s)
Mara Tableman, Eidgenoessische Technische Hochschule
Thomas P. Hettmansperger, The Pennsylvania State University


8. CONTRACT OR GRANT NUMBER(s)
N00014-80-C-0741

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR042-446

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistical and Probability Program Code 436
Arlington, VA 22217

12. REPORT DATE
June 1986

13. NUMBER OF PAGES
21

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Mann-Whitney-Wilcoxon test, nonparametric test, sign test, nonparametric confidence intervals, notched box plots.


20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
A two-sample test is studied which rejects the null hypothesis of equal population medians when two Wilcoxon distribution free confidence intervals are disjoint. A confidence interval for the difference in population medians is constructed by subtracting the endpoints of two one-sample confidence intervals. Two different ways to select the one-sample intervals are presented. A solution that specifies equal confidence coefficients for the one-sample intervals is recommended. All solutions are shown to have the same asymptotic (Pitman) efficiency as the Mann-Whitney two-sample test.

END
DTIC
8-86